First-Order Model Management Frameworks for Engineering Optimization

Natalia M. Alexandrov

NASA Langley Research Center

Slides accessible from http://fmad-www.larc.nasa.gov/mdob/MDOB

(better yet, search for ”MDOB”)

Outline

- **Introduction**
- **First-order model-management**
  - **Basic ideas**
  - **Example: Sℓ_1QP-AMMO framework**
  - **Computational example: variable-fidelity physics in AMMO**
- **Comments**

Introduction

- **Workshop's central question: How to endow modern large-scale PDE solvers with optimization capabilities?**
- **We ask: How to endow modern optimization solvers with large-scale PDE capabilities?**
  - **Theme of our talks:**
    - **An immediate solution of simulation-based optimization problems**
    - **Straightforward to integrate with simulations**
    - **Combine well-established engineering approximation concepts with the trust-region idea from nonlinear programming to ensure robust behavior**

Problem

- **The analysis or simulation problem: Given x, solve a system of coupled equations**

$$A(x, u(x)) = 0$$

for u that describes the physical behavior of the system.

- **The design problem (canonical formulation): Solve**

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x, u(x)) \\ \text{subject to} & c_i(x, u(x)) = 0, \;\; i \in \mathcal{E} \\ & c_i(x, u(x)) \leq 0, \;\; i \in \mathcal{I} \\ & x_l \leq x \leq x_u, \end{array}$$

where, given x, $u(x)$ is determined from $A(x, u(x)) = 0$.

- **In our context, "large-scale" means computationally expensive, regardless of the number of variables and constraints explicitly manipulated in optimization.**

Nature of the problem

- **Cost of solution is driven by simulations**
- **Number of variables and constraints explicitly used in optimization depends on problem formulation**
- **Focus on canonical formulation because need an immediate solution, while simulation codes are usually available as black boxes. Methodology applicable to other formulations.**
- **Function and derivative evaluation may not be robust**
- **Objective of work: reduce cost of optimization with simulations by minimizing the expense of using high-fidelity models in single-discipline optimization and MDO**

Ideas of 1st-Order Approx/Model Management Optimization (AMMO)

- **Derivative-based optimization techniques, including trust-region methods, rely on Taylor-series local models of objectives and constraints; variation in models - variation in derivative approximations**
- **AMMO takes the idea of local models further:**
  - **Use well-developed engineering approximation and modeling ideas**
  - **Replace Taylor-series models with general models that have trends similar to those obtained with high-fidelity models**
  - **Trust-region techniques provide global convergence**
- **Related work**
  - **Approximations long in use in engineering optimization (see refs in the paper)**
  - **Zero-order model management (comments later, time permitting)**

Ensuring local similarity of trends

Let $\tilde{f}$, $\tilde{c}_E$, and $\tilde{c}_I$ be some lower-fidelity models of f, c_E and c_I, respectively. At each major iteration k, x_k of an AMMO algorithm, the models are required to satisfy first-order consistency:

$$\tilde{f}(x_k) = f(x_k), \qquad \tilde{c}_E(x_k) = c_E(x_k), \qquad \tilde{c}_I(x_k) = c_I(x_k)$$

$$\nabla \tilde{f}(x_k) = \nabla f(x_k), \qquad \nabla \tilde{c}_E(x_k) = \nabla c_E(x_k), \qquad \nabla \tilde{c}_I(x_k) = \nabla c_I(x_k)$$

- **Models with this property locally mimic the behavior of first-order Taylor-series models around x_k**
- **Easily enforced when derivatives are available**

Enforcing First-Order Consistency

- **Multiplicative "β-correction", Chang, Haftka, Giles, Kao, 1993:**
  - **Given $\phi_{hi}(x)$ (say, f) and $\phi_{lo}(x)$, define $\beta(x) \equiv \frac{\phi_{hi}(x)}{\phi_{lo}(x)}$**
  - **Given x_k, build $\beta_k(x) = \beta(x_c) + \nabla\beta(x_k)^T(x - x_k)$**
  - **Then $\tilde{\phi}_k(x) = \beta_k(x)\phi_{lo}(x)$ satisfies the consistency conditions at x_k**
- **Additive correction (in the context of multigrid schemes), Lewis and Nash, 2000:**

$$\tilde{\phi}_k(x) = \phi_{lo}(x) + [\phi_{hi}(x_k) - \phi_{lo}(x_k)] + [\nabla\phi_{hi}(x_k) - \nabla\phi_{lo}(x_k)]^T(x - x_k)$$

Examples of Variable-Fidelity Models for Use in AMMO

- **Data-fitting models (polynomial RS, splines, kriging)**
  - **Rely directly on hi-fi information; do not require derivatives; simple to construct; difficult to sample; "curse of dimensionality"**
- **Reduced-order models**
  - **Use reduced-order bases (constructed as a span of solutions and possibly derivatives at some points) to represent field variables at other points**
- **Variable-accuracy models**
  - **Converge analyses to a user-specified tolerance**
- **Variable-resolution models**
  - **Executing a single physical model on meshes of varying degree of refinement**
- **Variable-fidelity physics models**
  - **E.g., in aerodynamics, physical models range from inviscid, irrotational, incompressible flow to Navier-Stokes equations for nonlinear viscous flow**

Convergence vs. Performance

- **Convergence analysis relies on the consistency conditions and standard assumptions for the convergence analysis of the underlying algorithm (see paper for three examples)**
- **For convergence, need only a notion of two models, one arbitrarily designated “high fidelity” or “truth”, the other - “low fidelity”**
- **Practical efficiency**
  - **Problem/model dependent**
  - **Depends on the ability to transfer computational load onto low-fidelity computation, which...**
  - **Depends on the predictive quality of the low-fidelity models (surrogates)**
  - **In the worst case, AMMO is conventional optimization**

Example: AMMO Based on Sℓ_1QP

- AMMO can be used with any derivative-based algorithm; to date, implemented and tested AMMO based on five algorithms
- **Principle:** a simple implementation with maximum use or existing software
- **Problem:** have not found software suitable for simulation-driven optimization
- **Resolution:** writing our own
- **Meanwhile:** nonsmooth exact penalty functions - a potential alternative to SQP; simple merit function, similar convergence properties (Fletcher 1989)

Consider a composite penalty function

$$\mathcal{P}(x; h) \equiv f(x) + h(c(x)),$$

where f and c are smooth and h is convex but possibly only continuous.

Sℓ_1QP

Fletcher's choice of $\mathcal{P}$ is the penalty function

$$\mathcal{P}(x;\sigma) = f(x) + \sigma \sum_{i\in E} |c_i(x)| + \sigma \sum_{i\in I} \max\{0, c_i(x)\}.$$

This is an exact penalty function if σ satisfies

$$\sigma > \min_{i\in L} |\lambda_i|,$$

where L is the set of all multipliers for the NLP. The model of $\mathcal{P}$ is

$$m(x_k, s;\sigma) \equiv q(x_k, s) + \sigma \sum_{i\in E} |l_i(x_k, s)| + \sigma \sum_{i\in I} \max\{0, l_i(x_k, s)\},$$

where $q(x_k, s)$ is the quadratic model of f and $l_i(x_k, s)$ are linearizations of constraints. The prototype Sℓ_1QP finds global solutions s_k of

$$\begin{array}{ll} \underset{s}{\textbf{minimize}} & m(x_k, s;\sigma) \\ \textbf{subject to} & \| s \|_\infty \le \Delta_k \end{array}$$

Sℓ_1QP, continued

The step is evaluated by examining

$$\rho_k = \frac{\mathcal{P}(x_k;\sigma_k) - \mathcal{P}(x_k + s_k;\sigma_k)}{m(x_k, 0;\sigma_k) - m(x_k, s_k;\sigma_k)} \text{ as follows:}$$

Select $0 < r_1 < r_2 \leq 1$ **and** $0 < \kappa_1 < 1 < \kappa_2$.
Typical values are $r_1 = 0.25$, $r_2 = 0.75$, $\kappa_1 = 0.25$, $\kappa_2 = 2$.

Set $x_{k+1} = \begin{cases} x_k & \text{if } \rho_k \leq 0 \\ x_k + s_k & \text{otherwise.} \end{cases}$

Set $\Delta_k = \begin{cases} \kappa_1 \| s_k \| & \text{if } \rho_k < r_1 \\ \kappa_2 \Delta_k & \text{if } \rho_k > r_2 \text{ and } \| s_k \| = \Delta_k \\ \Delta_k & \text{otherwise.} \end{cases}$

Sℓ_1QP-AMMO Model and Algorithm

$$m(k, x_k, s; \sigma) \equiv \tilde{f}(k, x_k, s) + \sigma \sum_{i \in E} |\tilde{c}_{E,i}(k, x_k, s)| + \sigma \sum_{i \in I} \max\{0, \tilde{c}_{I,i}(k, x_k, s)\}$$

whose components satisfy the consistency conditions. Note that the model m depends on k. as follows.

Initialization: Choose x_0, Δ_0, and constants as above.

Do $k = 0, 1, \ldots$ until convergence:

Model construction:

Construct model $m(k, x_k, s; \sigma_k)$ of $\mathcal{P}$

Step computation:

Solve for s_k $\begin{cases} \underset{s}{\text{minimize}} & m(k, x_k, s; \sigma_k) \\ \text{subject to} & \| s \| \leq \Delta_k \end{cases}$

Step evaluation: Compute ρ_k. Accept or reject the step based on ρ_k as above.

Updates: Update x_k, Δ_k based on ρ_k as above.

End do

Convergence of Sℓ_1QP-AMMO

Theorem:

Let $f, c_E, c_I \in C^2(\Omega)$ have bounded second derivatives on a bounded $\Omega \subset I\!R^n$. Let $\tilde{f}, \tilde{c}_E, \tilde{c}_I \in C^2(\Omega)$ be any models of f c_E, and c_I, respectively, that satisfy the first order consistency conditions and have uniformly bounded second derivatives on Ω. Let $\{x_k\} \in \Omega$ be the sequence of iterates generated by Sℓ_1QP–AMMO. The there exists an accumulation point x_* at which the first-order optimality conditions for minimizing $\mathcal{P}$ hold, that is,

$$\underset{\lambda \in \partial h_*}{\text{maximize}} \; (g_* + \nabla c_* \lambda)^T s \geq 0 \text{ for all } s,$$

where ∂h_* is the generalized derivative of h.

An Alternative Sℓ_1QP-AMMO

Impose the following conditions on the model and the trial step:

- **Smoothness:** The model m is locally Lipschitz continuous and regular with respect to s for all (x, σ) and continuous in (x, σ) for all s.
- **Zero-order matching:** The values of the function and model coincide when $s = 0$.
- **First-order matching:** The generalized directional derivatives of the function and model coincide when $s = 0$.
- **Bounded parameters:** The set of problem parameters is closed and bounded.
- **Sufficient decrease:** For any x_*, there exist constants $\delta, \epsilon, \kappa \in (0,1)$ such that s_k satisfies

$$m(k, x_k, 0, \sigma_k) - m(k, x_k, s_k, \sigma_k) \geq \kappa \parallel g(x_k) \parallel \min\{\delta, \Delta_k\},$$

where $g = \arg\min_{g \in \partial f} \parallel g \parallel$. These conditions are summarized in CGT 2000.

An Alternative Sℓ_1QP-AMMO, continued

In Sℓ_1QP-AMMO, the smoothness, boundedness, zero- and first-order matching conditions are satisfied by assumption. Guaranteeing sufficient decrease - in progress.

Updates for Sℓ_1QP-AMMO with sufficient decrease

Select $\Delta_{max} > 0, 0 < r_1 \le r_2 \le 1$ **and** $0 < 1/\kappa_3 \le \kappa_1 \le \kappa_2 < 1 < \kappa_3$.

Set $(x_{k+1}) = \begin{cases} x_k + s_k & \text{if } \rho_k \ge r_1 \\ x_k & \text{otherwise.} \end{cases}$

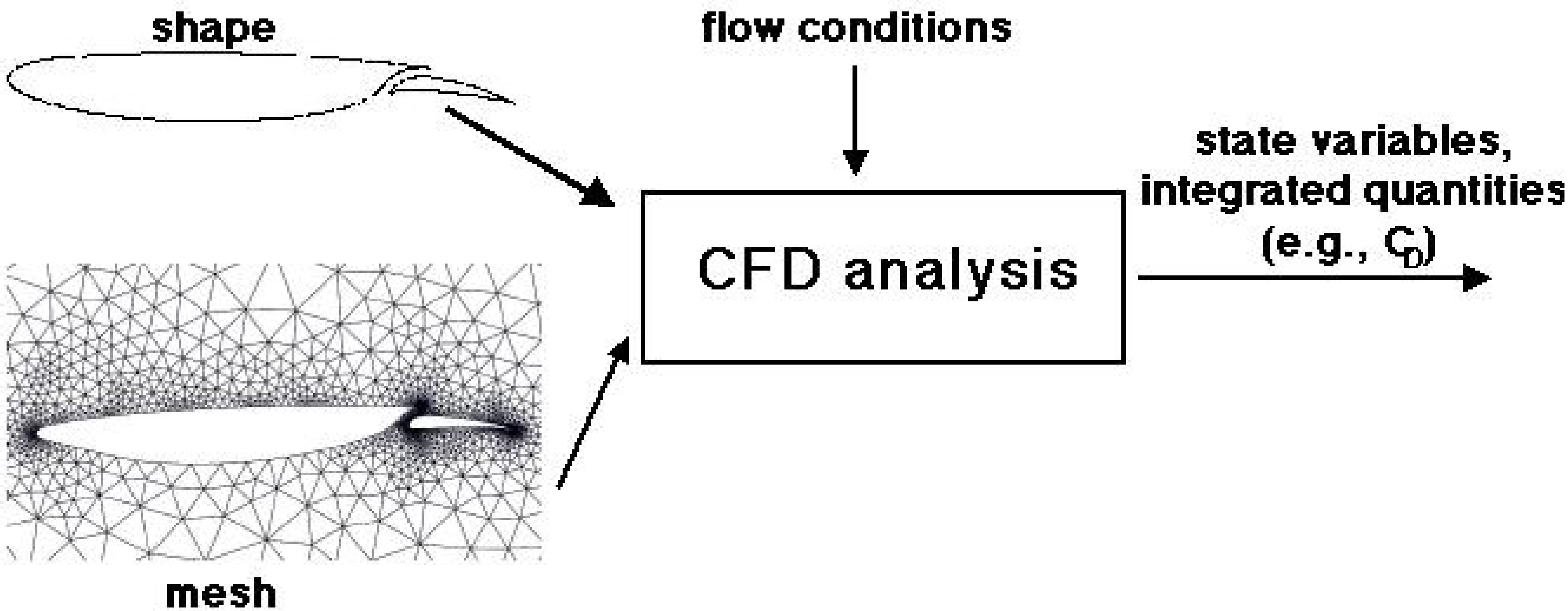
Set $\Delta_{k+1} \in \begin{cases} [\kappa_1 \Delta_k, \kappa_2 \Delta_k] & \text{if } \rho_k < r_1 \\ [\kappa_2 \Delta_k, \Delta_k] & \text{if } \rho_k \in [r_1, r_2) \\ [\kappa_3 \Delta_k, \kappa_2 \Delta_{max}] & \text{if } \rho_k \ge r_2. \end{cases}$

Convergence to a first-order critical point is immediate under these conditions (see, e.g., Theorem 11.2.5 in CGT 2000).

Computational Demonstrations

- **Because of data-fitting model limitations, we have focused on models that are independent of the number of variables**
- **Independence wrt dimension is important: in preliminary design, problems of modest size number O(100) variables**
- **AMMO admits a wide variety of models and algorithms; demonstrations are aimed at accumulating realistic experience to validate the algorithmic performance**
- **Because we cannot predict *a priori* the relative descent characteristics of models, must include cases of favorable and unfavorable relationship between models**
- **Aerodynamic shape optimization is a good test problem: practically important, computationally intensive, comes in a variety of dimensions**

Demonstration Problems: Aerodynamic Optimization



minimize Integrated quantities, such as $-\frac{L}{D}$ $\left(\frac{\text{lift}}{\text{drag}}\right)$ or C_D (drag coefficient)

subject to constraints on, e.g., pitching and rolling moment coefficients, etc.

$x_l \leq x \leq x_u$

Managing Variable-Fidelity Physics Models: Multi-Element Airfoil

(AIAA-2000-4886, Alexandrov, Nielsen, Lewis, Anderson)

- **A two-element airfoil designed to operate in a transonic regime — inclusion of viscous effects is very important**
- **Governing equations: time-dependent Reynolds-averaged Navier-Stokes**

$$A\frac{\partial Q}{\partial t} + \oint_{\partial\Omega} \vec{F}_i \cdot \hat{n} dl - \oint_{\partial\Omega} \vec{F}_v \cdot \hat{n} dl = 0,$$

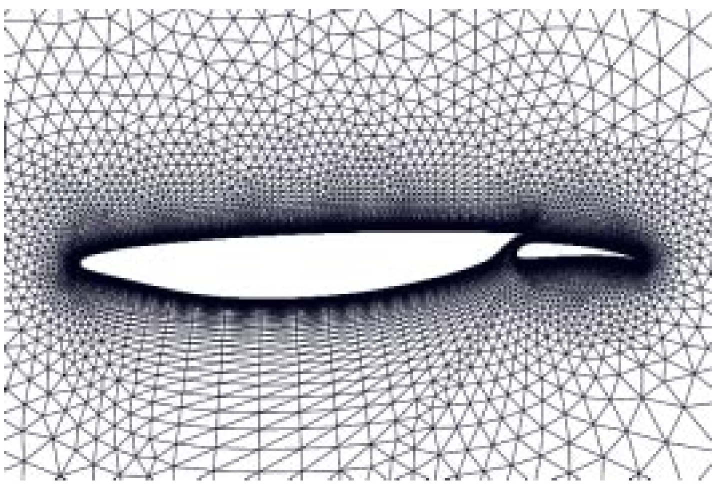
  where $\vec{F}_i$ and $\vec{F}_v$ are the inviscid and viscous fluxes, respectively

- **Flow solver (FUN2D) – unstructured mesh methodology (Anderson, 1994)**
- **Sensitivity derivatives – hand-coded adjoint approach (Anderson, 1997)**
- **Conditions:**
  - $M_\infty = 0.75$
  - $Re = 9 \times 10^6$
  - $\alpha = 1^\circ$ **(global angle of attack)**

Multi-Element Airfoil, cont.

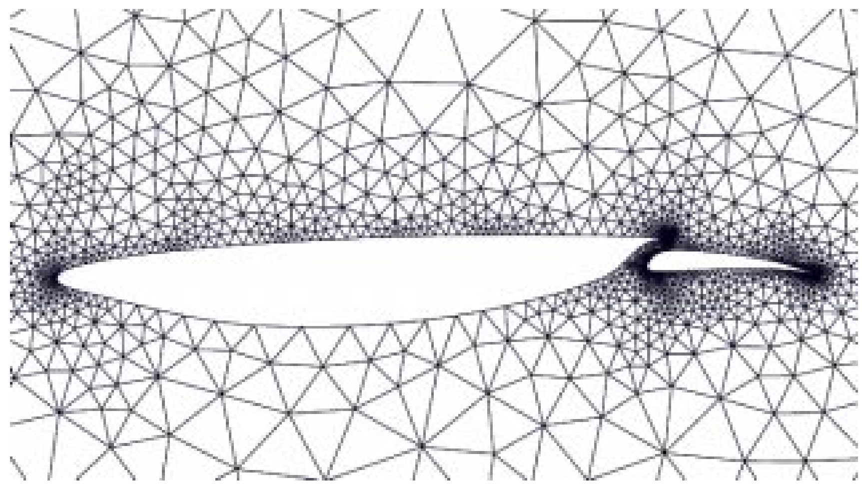
- **Hi-fi model – FUN2D analysis in RANS mode**
- **Lo-fi model – FUN2D analysis in Euler mode**
- **Computing on SGI Origin^{TM} 2000, 4 R1OK processors**

Viscous mesh:
10449 nodes and 20900 triangles



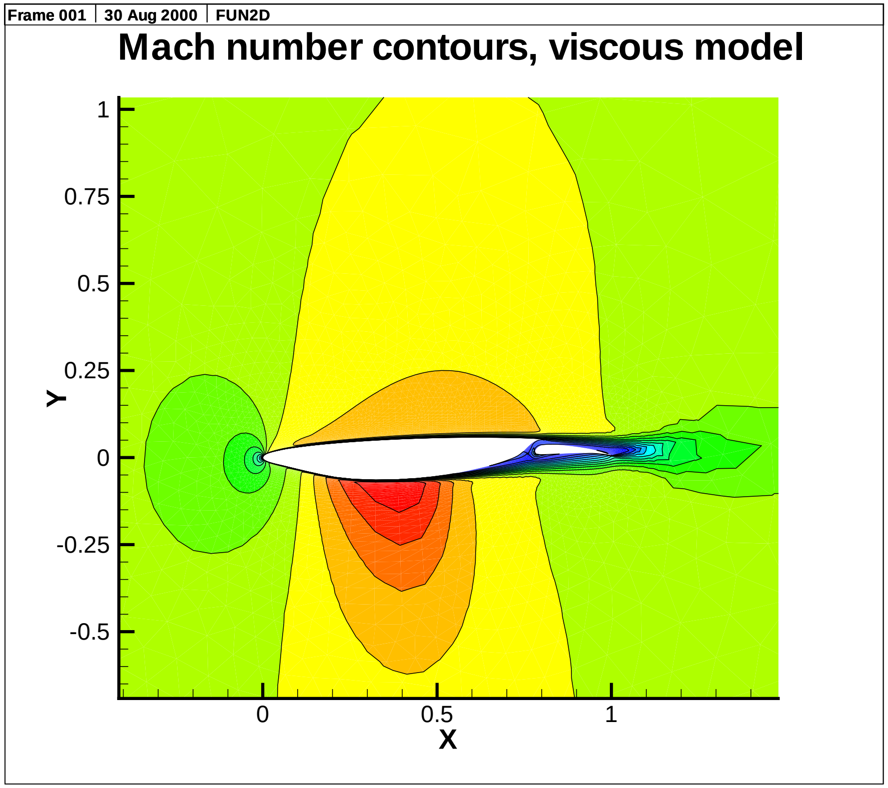
t/analysis $\approx$ 21 min
t/sensitivity $\approx$ 21 or 42 min

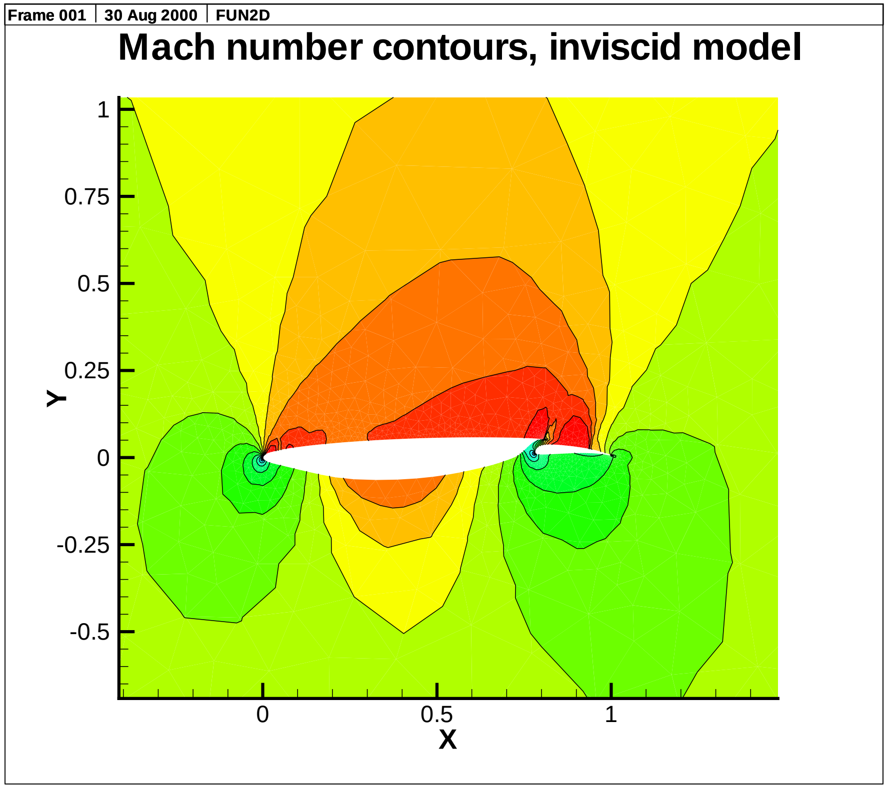
Inviscid mesh:
1947 nodes and 3896 triangles



t/analysis $\approx$ 23 sec
t/sensitivity $\approx$ 100 or 77 sec

Multi-Element Airfoil: Viscous Effects





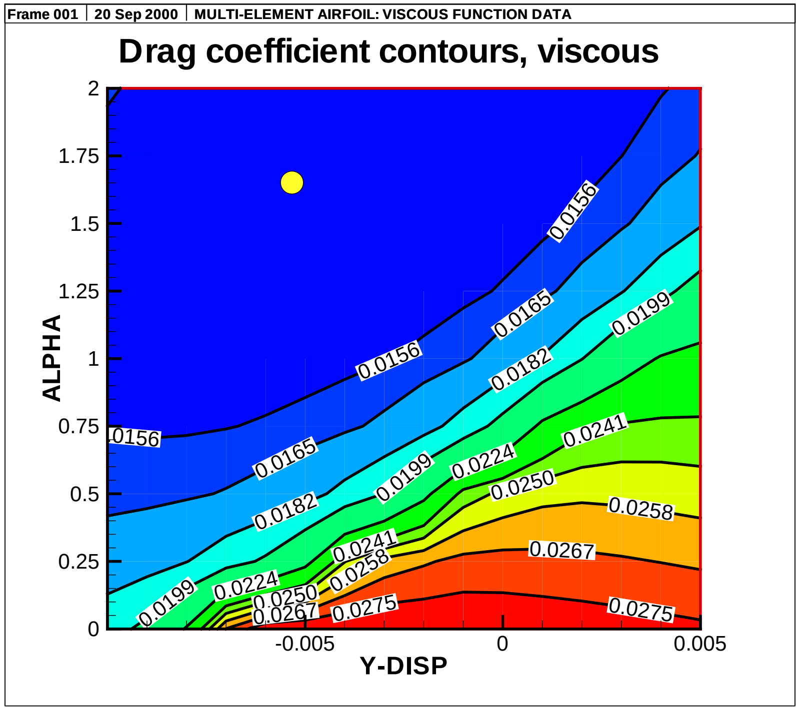
- **Boundary and shear layers are visible in the viscous case.**

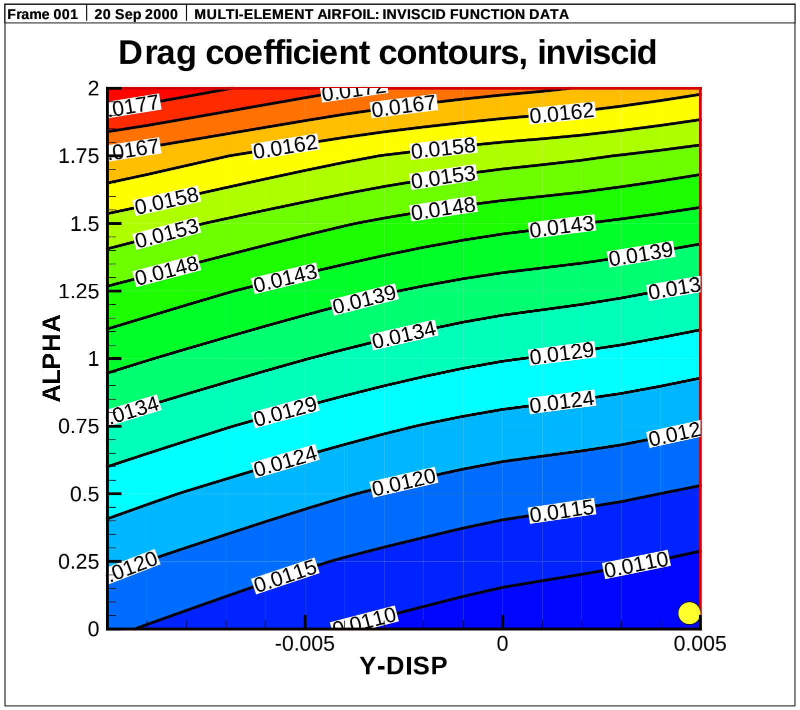
Multi-Element Airfoil: Computational Experiments

- **Objective function:** minimize drag coefficient subject to bounds on variables
- **Case 1:** (for visualization)
  - Variables: angle of attack, y-displacement of the flap
  - Solve problem with hi-fi models alone using a commercial optimization code (PORT, Bell Labs)
  - Solve the problem with AMMO, PORT used for lo-fi subproblems
- **Case 2:**
  - Variables: angle of attack, y-displacement of the flap, geometry description of the airfoil; 84 variables total
  - Same experiment

Multi-Element Airfoil: Models

- **Time/function for inviscid model negligible compared to viscous model**
- **Descent trends are reversed — unusual but a good test**

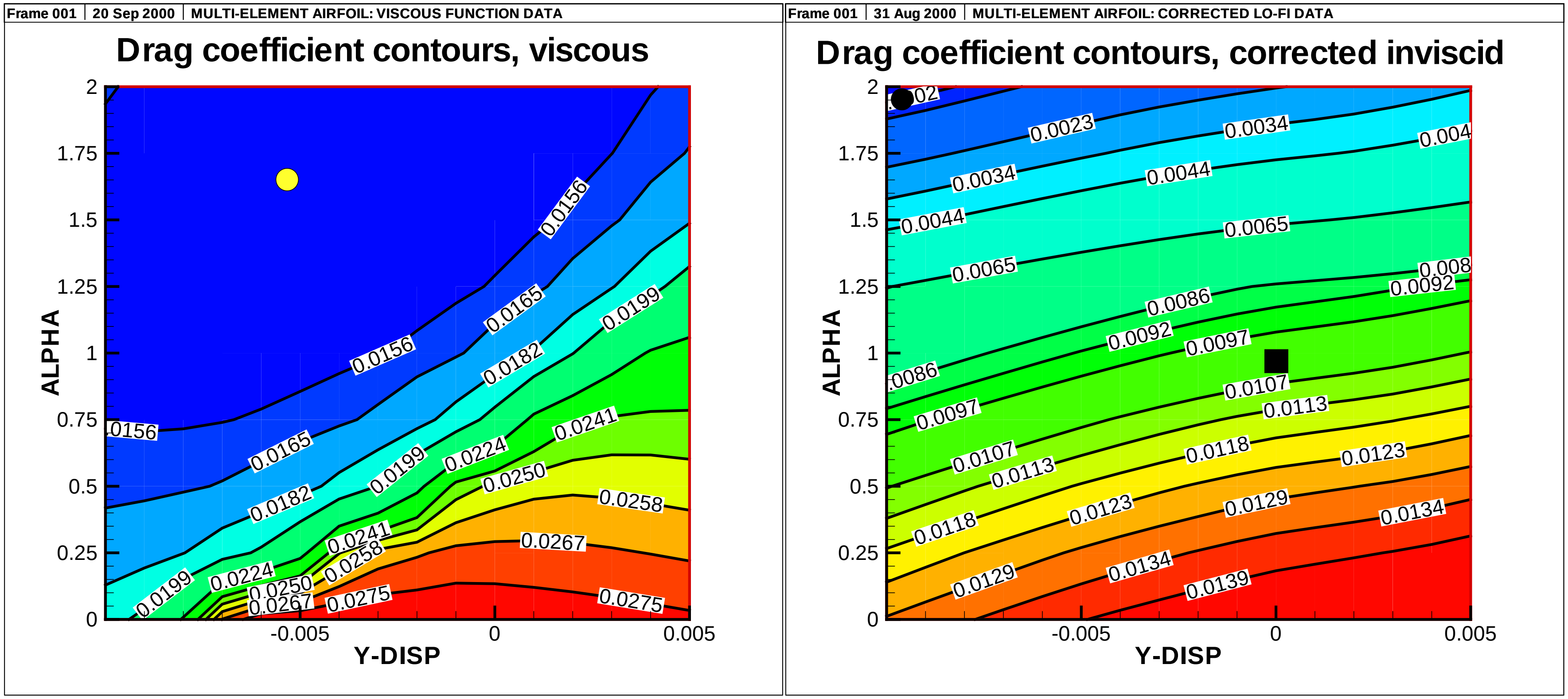




Multi-Element Airfoil: AMMO Iterations with 2 Variables

Iteration 1. Starting point: $\alpha = 1.0$, y-disp $= 0.0$

High-fidelity objective vs. corrected low-fidelity objective



New point: $\alpha = 2.0$, y-disp $= -0.01$

Multi-Element Airfoil: AMMO Iterations with 2 Variables, cont.

- **Similar effect in the next iteration**
- **Solution ($\alpha = 1.6305°$, flap y-displacement $= -0.0048$) located at iteration 2**
- $C_D^{\text{initial}} = 0.0171$ **at ($\alpha = 1°$, flap y-displacement $= 0$)**
- $C_D^{\text{final}} = 0.0148$**, a decrease of approximately** 13.45%**.**

Multi-Element Airfoil: Performance Summary

Notation: No. functions / No. Gradients

Test	hi-fi eval	lo-fi eval	total t	factor
PORT with hi-fi analyses, 2 var	14/13		$\approx$ 12 hrs	
AMMO, 2 var	3/3	19/9	$\approx$ 2.41hrs	$\approx$ 5
PORT with hi-fi analyses, 84 var	19/19		$\approx$ 35 hrs	
AMMO, 84 var	4/4	23/8	$\approx$ 7.2hrs	$\approx$ 5

Concluding Remarks

- **AMMO techniques can be used in conjunction with any derivative-based method**
- **Sensitivity information is becoming increasingly available with analysis codes (automatic differentiation, efficient adjoint techniques) — we believe first-order AMMO will prove helpful**
- **The results are promising**
  - **Relatively straightforward implementation and integration with simulations**
  - **Convergence analysis is a direct consequence of the analysis for underlying optimization algorithms**
  - **Significant savings over conventional optimization in terms of hi-fi evaluations**
  - **Capture descent behavior with the help of corrections, despite sometimes significant dissimilarities between lo-fi and hi-fi models**
  - **First-order information is indispensable when model trends are dissimilar**